CENTRALIZATION AND AUTOMATION OF CANCER EXPERIENCE*

FRED G. CONRAD, Col. USAF, MC

Director, Armed Forces Central Medical Registry

ROBERT G. ROSSING, M.D., AND HENRY R. BALES, JR.

School of Aerospace Medicine
Brooks Air Force Base
San Antonio, Texas

INTRODUCTION

THE primary and, possibly, the only advantage to the centralization of tumor registries is to accumulate large numbers of cases so that clinical studies may gain statistical validity. All biological studies deal with a number of variables which must be defined and compared in order to make the study worthwhile. Each variable will diminish the number of cases available for analysis, thereby decreasing the opportunity to draw a valid conclusion. Few hospitals and only a rare physician will have sufficient experience with a single disease treated in multiple ways so that contention concerning methods of treatment may be resolved. This is why some arguments over therapeutic modalities have been waged for 70 years without resolution. It is now possible, through centralization of experience from several sources and through the use of automated systems, to file abstracts of many thousands of cancer records and to compare large numbers of clinical and therapeutic variables rapidly. This affords a statistically valid comparison of treatment modalities which may then be applied for the benefit of the individual patient. We shall describe a few problems associated with the development of a centralized tumor registry and our solutions to these problems.

The initial consideration is to determine the purpose of the registry. Is it to be an epidemiologic and demographic registry, or primarily a registry dealing with clinical and therapeutic information? We chose the latter and established as our purpose "to improve the care of cancer

*Presented at the 26th Annual Meeting of the Society of Medical Consultants to the Armed Forces held in Washington, D.C., November 22, 1971.

patients through analysis of treatment results." From this one may decide upon the output or automated retrieval, which will meet the requirements stated in the purpose. The output must include precise diagnosis, definition of the extent of the tumor, therapeutic approach, status of the patient, and quality of survival, recurrence, and secondary therapy. These may be stated as a series of questions: What is the tumor? How extensive is it? How does it affect the patient? How is it treated? To what extent has the patient been helped and to what extent debilitated by treatment? Did the tumor recur? If so, where, when, and what further was done? What is the ultimate result?

Input Data

By knowing the questions we want answered it is relatively simple to determine the input of data necessary to accomplish the desired retrieval. At this point it becomes essential to work with a statistician and computer programmer in order to be certain that the input will be sufficiently objective to yield statistically valid output and will be in a coded format compatible with the computer to be used.

The input may be divided into three broad areas, including basic identification data, initial clinical data, and updated data. Basic information includes the patient's name, social security number, sex, race, date of birth, accession number, hospital of diagnosis, and physician. Initial clinical data include site, histology, date of diagnosis, duration of symptoms, TNM stage, clinical stage, tumor size, basis for diagnosis, and primary treatment. Supplementary up-to-date information, which is obtained every six months for two years and annually thereafter, always includes present status and quality of survival and may include date of recurrence, secondary, tertiary, and quaternary therapy, date of death, autopsy, metastatic sites of involvement, and additional primaries.

It is necessary to describe briefly some of the more important items of input and the reasons behind some of our decisions. We felt that it was essential to have a very detailed listing of site and histologic diagnosis. One often hears of the relative value of "splitters" and "lumpers" in medicine, but a registry must adhere to the principle of "splitting." The abstracted information may always be "lumped" by automated means later, but it may be "split" further at a later time only by review of each individual record; this completely destroys the value of an automated retrieval system. Our system is based on the Systematized

Nomenclature of Pathology (SNOP) for sites and histology, with a few modifications including Rappaport's classification for lymphosarcoma,[1] Lukes' for Hodgkin's disease,[2] and that of Dixon and Moore for testicular tumors.[3] The tumor is staged pathologically by means of the T (tumor), N (lymph node), M (metastases) systems.[4] This is more complex than what one usually requires and it may be that the broad terms recommended by the American College of Surgeons such as in situ, local, regional, and disseminated would suffice. However we felt that it was important to differentiate the primary tumor, so that the various pathologic stages of invasiveness could be included and still maintain the differentiation from lymph-node status such as clinically uninvolved, pathologically uninvolved or, if involved, the level and degree of involvement. If it is important in evaluating results to know if a melanoma is in stage II and not stage III, or if there should be a difference in therapeutic approach when first-level lymph nodes are positive as compared to third-level involvement, then broad categorization will yield inadequate output and will destroy the ability of the registry to respond to special questions. Since we feel that these clinical criteria are important, we take the time to ascertain the exact pathologic extent of the disease. It is also important, especially in cancer of the female genital organs, to record the clinical stage of disease, since this is the primary determinant of therapy. Therefore clinical staging is included as a separate entry and is used primarily for cancer of the cervix and Hodgkin's disease. If a celiotomy is performed, the findings are recorded under pathologic staging. Sizes of tumors are recorded whenever possible by listing the largest diameter, measured in millimeters.

Therapy is recorded as a four-part entry, which includes surgery, radiation, chemotherapy, and hormone therapy. Primary surgery includes biopsy, local resection, wide resection, lymphadenectomy; this scheme attempts to consider all possible modes of primary surgical approach, and it codes only the most definitive. Radiation therapy includes preoperative radiation, postoperative, therapeutic, palliative, radioisotope, etc. Chemotherapy is less detailed, using broad categories of drugs such as alkylating agents, antimetabolites, perfusion, combination chemotherapy, etc., but it is possible to code special studies such as those on COMP, MOPP, CO-AP, etc. Hormone therapy includes estrogen, testosterone, corticosteroids, thyroid, etc. Data about all

four modalities of treatment must be entered in each record. It is possible to enter four different courses of therapy during the patient's lifetime but unless recurrence takes place, only the primary therapy would be entered. The others are reserved for treatment after recurrence.

Exact dates are required for date of birth, diagnosis, recurrence, and death; this allows easy determination of age at any particular time and also calculation of means and medians in terms of months.

We feel that it is important not only to know absolute survival, but also quality of survival. To what extent is a patient functional or symptom-free? Since the information is updated only once a year and variations may occur during that time, it is necessary to be rather broad in this category. Hence we use: "fully functional," "symptomatic but functional," "symptomatic to an extent that interferes with full activity," and "severely limited or bedridden." It is also important to determine if the patient is limited by the therapy: e.g., by hindquarter amputation and inability to adjust to a prosthesis, or limited by recurrent disease.

Abstracting Tumor Records

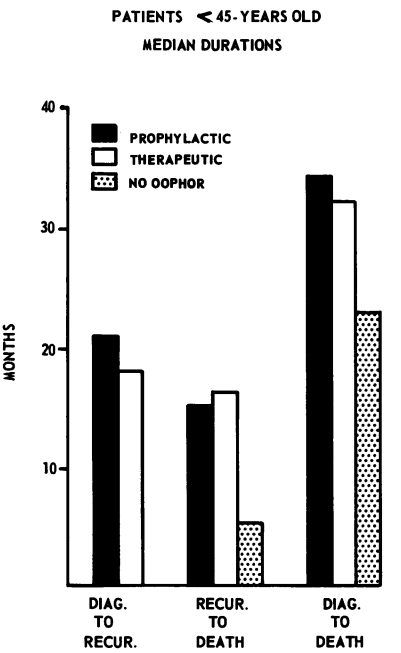
The abstracting and coding of input information should be done by a central core of trained persons; not by the registrar at each institution. This is probably just a personal prejudice, but we feel strongly that the director should have frequent contact with the coders, who in turn should have immediate access to the director in order to resolve even the smallest questions. No matter how detailed a syllabus of coding instructions is written, subjective and quantitative decisions will always be necessary. Great efforts should be made to keep these as standardized as possible to avoid arbitrary decisions. If a central registry is consolidating the records of a single city, it may be considered expedient for the coders to visit each hospital periodically and code the records locally. Again we object to this for several reasons. Quite often all the information required to abstract a record may not be immediately available. If such a record is centrally located it may be placed in a "hold" file and the additional information requested. This takes very little effort for the coder, but if the same record is held locally the coder may be tempted to abstract the incomplete record. Also, not having immediate access to the director or coding supervisor may lead

to an increased number of arbitrary decisions. Since it is important that a complete copy or microfilm of the tumor record be available at the central registry, this too would obviate the necessity to code each record locally.

The reason for maintaining a complete record centrally is that a certain number of clinical studies, perhaps 5%, will require more detailed or different information than that which has been coded. In order to comply with such requests we have microfilmed all records and can make a "microfiche" copy of each record at a cost of 3.5 cents per record, e.g., $22.75 and two and one half hours of time to reproduce the records of 650 patients with cancer of the thyroid. These records may then be given to the requesting physicians along with the loan of a microfilm reader. He may then study the records at his leisure and destroy them when he is finished. This allows him to study the entire central registry experience, rather than merely the records of a single hospital.

Personnel

It is our opinion that the director should be a physician, preferably a clinician knowledgeable in oncology, so that the direction of the registry will remain clinically oriented and requests for information may be expertly interpreted to the programmer. The number of persons required to operate a registry depends upon several factors. It is assumed that a central core of persons will abstract and code all the charts. If follow-up letters and contact devolve upon the registrar of each participating hospital, a coding clerk will be able to register, code, update, and maintain approximately 1,500 records per year with information of the kind that we have described. If less information is required more records may be registered, and if a great deal of detailed information is required fewer records will be processed. If the coding clerk is required to maintain follow-up of all or most of the patients registered, as we do, each coding clerk will register only approximately 800 patients per year. Our experience has been that every 1,000 hospital beds will generate approximately 650 new cancer patients per year. Therefore, it is quite easy to determine the number of coding clerks required by knowing the number of hospital beds involved in the centralized registry. The actual mechanics of the automation processes, which are extremely complicated in order to ensure that information submitted is



A comparison of patients undergoing prophylactic oophorectomy versus therapeutic oophorectomy shows no significant difference in the median duration from diagnosis to recurrence or the total period from diagnosis to death.

logical, is free from coding errors, and is not duplicated, are not discussed in the present statement, but a detailed description of these processes and the integration of updated or changed information has been published as a USAF Technical Document[5] and is available on request.

Clinical Studies

It might be well to demonstrate some of the ways that the registry has been used since it originated four years ago in order to demonstrate the types of clinical studies which may be accomplished easily and rapidly.

Cancer of the breast. There has been much controversy regarding prophylactic oophorectomy in patients who have carcinoma of the breast and metastases in axillary lymph nodes. Kennedy[6] demonstrated that the mean duration of the recurrent-free interval is longer in patients who have had prophylactic oophorectomy, compared with patients in whom the oophorectomy was done for therapeutic purposes.

TABLE I. BREAST IRRADIATION: COMPARISON OF RECURRENCE AND SURVIVAL

	No. postop. irradiation	*Postop. irradiation*	*Chi square*
Total patients	132	608	—
Number recurrence	64 (48%)	280 (45%)	N.s.*
Number dead	47 (36%)	214 (35%)	N.s.
Number local recurrence	30 (23%)	94 (15%)	N.s.
Median time to recurrence	18 Mo.	19 Mo.	N.s.
Median time of survival	31 Mo.	32 Mo.	N.s.

*N.s. = not significant.

Kennedy used this as the rationale to advocate the prophylactic procedure. Unfortunately, in a study of this type the median is the more important factor, not the mean. The reason for this is that after a definitive operation recurrence may develop almost immediately or many years later. If the mean time for recurrence is 20 months, a patient who suffers recurrence at 100 months will have five times the weight of a patient who has recurrence at one month. This may lead to an erroneous conclusion, since each patient should contribute equal weight to the result. Therefore only median results should be considered. Such a study has been done using our data and the result is shown in the accompanying figure, which shows no difference in the median duration of recurrence-free interval or total survival. The value of determining quality of survival was also demonstrated in this study, because even though the patients who underwent therapeutic oophorectomy did not survive any longer than those patients who had prophylactic oophorectomy, their quality of survival was much improved during the period from the development of recurrence until death.

A study of prophylactic postoperative irradiation in this group of patients shows no decrease in recurrence and no difference in the rate of recurrence or survival. There is a decrease in the incidence of recurrence in the chest wall, 15% versus 23%, but the difference is not statistically significant at the 0.05 level (Table I). Therefore we see no rationale for subjecting a patient to four weeks of cobalt therapy, and we reserve irradiation until recurrences become evident in the chest wall or elsewhere.

TABLE II. THE HAZARD RATE OF RECURRENCE IN PATIENTS WITH MALIGNANT MELANOMA*

Site	*Lymph nodes*	*Year risk* $\leq 5\%$
Head-neck	Postitive	†
	Negative	3
Trunk	Postitive	7
	Negative	6
Extremities	Postitive	4
	Negative	1

*Reproduced by permission from: Conrad, F. G. et al.: The hazard rate of recurrence in patients with malignant melanoma. *Aero. Med. 42:*1219-25, 1971.
†Insufficient cases upon which to base an estimate.

Chronic granulocytic leukemia. For years initial remission in this disease has been achieved equally well with Busulfan or splenic irradiation. It is only recently that a British Cooperative Study[7] and also our results, have revealed harmful long-term effects from the use of radiation. Irradiated patients not only pass into a "blast" phase of their disease sooner and have a median survival almost two years shorter than those treated with Busulfan, but the irradiation also precludes long-term survival. No patient treated with radiation survived longer than seven years whereas 15 Busulfan-treated patients have done this, and two are living beyond 12 years from the time of diagnosis.[8]

Malignant melanoma. Some studies require determination of means or medians using the absolute number of months between events, but for the calculation of life tables[9] it is necessary to know the status of a group of patients at specific times such as six months, five years, etc. It is also necessary to know whether patients are living with or without recurrent disease at these times. Hence the coding system should include the following as part of the patient's status: alive without evidence of disease, alive with recurrence, dead of cancer, dead of other disease, lost to follow-up, etc. In this way, life tables may be programmed and retrieved with little or no additional effort. Life tables are absolutely essential for comparing results, whether the comparisons are made between modes of therapy, age groups, sex, site of primary, or even contributing hospital. Life tables yield relatively smooth descending curves of per cent survival, either absolute or recurrence–free, at each point in time. Their major deficiency is that the curves do

TABLE III. INCIDENCE OF RECURRENCE IN PATIENTS FOLLOWED AT LEAST FIVE YEARS WITH LYMPH NODES "NEGATIVE"

Site	*Wide resection* Number	Percent	*Lymphadenectomy* Number	Percent	*Chi square*
Head-neck	61 (19)	31	26 (10)	38	N.s.
Trunk	115 (53)	46	29 (13)	45	N.s.
Extremity	83 (21)	25	43 (7)	16	N.s.
Total	259 (93)	36	98 (30)	31	N.s.

*Reproduced by permission from: Conrad, F. G.: Treatment of malignant melanoma: Wide excision alone versus lymphadenectomy. *Arch. Surg. 104*:587-93, 1971.

N.s. = not significant.

The parentheses enclose the absolute numbers of recurrence.

not indicate clearly the peak periods of recurrence. Further programs, as developed by Gehan,[10] make it possible to calculate those periods of time during which a patient may be especially apt to develop a recurrence. This is the so-called "hazard rate"; we find it most important in dealing with flying personnel. Calculation of hazard rates and an acceptance of less than one chance in 20 (a 5% hazard rate) that a patient entering a particular time interval will develop a recurrence during that interval has allowed us to modify previous arbitrary standards as to when a pilot may return to the cockpit following an operation for cancer. Table II indicates the time periods as related to malignant melanoma of the skin.[11]

Comparison of different modes of surgical operation, such as wide resection versus lymphadenectomy, is possible by means of life tables. An evaluation of 640 patients with malignant melanoma and construction of recurrence-free life tables has shown no benefit in patients treated with lymphadenectomy over those who were treated with wide resection alone (Table III).[12]

These are but a few of the many studies that have been accomplished during the past few years. It is fascinating to think that studies which would have required many hundreds of man-hours just a few years ago may now be done in a few minutes, and can be based upon far greater numbers of cases than were ever available to a single physician. Because of limited programmer time, we are limited to performing three to five special studies per month. However this service is available

not only to military physicians but to any physician interested in analysis of a clinical cancer problem.

SUMMARY

A brief description of the function and format of the newly formed Armed Forces Central Medical Registry is presented. The registry is clinically oriented to answer requests concerned with clinical and therapeutic cancer problems. The reasons for specific data input are discussed and several clinical problems of therapy are presented to demonstrate the types of information which are automated and easily retrievable.

REFERENCES

1. Rappaport, H.: Tumors of the Hematopoietic System. In: *Atlas of Tumor Pathology,* Sec. III, Fasc 8. Washington, D.C., Armed Forces Institute of Pathology, 1966.
2. Lukes, R. J.: The Pathologic Picture of the Malignant Lymphomas. In: *Proc. Internat. Conf. Leuk-Lymph.* Philadelphia, Lea & Febiger, 1968.
3. Dixon, F. J., Moore, R.: Tumors of the Male Sex Organs. In: Atlas of *Tumor Pathology,* 1952, VIII., Fasc 31 and 32. Washington, D.C., AFIP, 1966.
4. International Union Against Cancer (UICC): *TNM Classification of Malignant Tumors.* Geneva, de Buren, 1968.
5. Bales, H. R., Jr., Rossing, R. G. and Conrad, F. G.: *The Air Force Central Tumor Registry: A Description of Format, Automation Processes and Programming Schemas.* USAF Tech. Doc., AMD, Brooks Air Force Base. In press.
6. Kennedy, B. J., Mielke, P. W., Jr., and Fortuny, I. E.: Therapeutic castration versus prophylactic castration in breast cancer, *Surg. Gynec. Obstet. 118*:524-40; 1964.
7. Med. Res. Counc. Rep.: Chronic granulocytic leukemia: Comparison of radiotherapy and Busulfan therapy. *Brit. Med. J. 1*:201-08, 1968.
8. Conrad, F. G. Survival in chronic granulocytic leukemia: Splenic irradiation versus Busulfan. *Arch Intern. Med.* In press.
9. Cutler, S. J., and Ederer, F.: Maximum utilization of the life table method in analyzing survival. *J. Chron. Dis. 8:* 699-712, 1958.
10. Gehan, E. A.: Estimating survival functions from the life table. *J. Chron. Dis. 21*:629-42, 1969.
11. Conrad, F. G., Rossing, R. G., Allen, M. F. and Bales, H. R., Jr.: The hazard rate of recurrence in patients with malignant melanoma. *Aero. Med. 42:* 1219-25, 1971.
12. Conrad, F. G.: Treatment of malignant melanoma: Wide excision alone versus lymphadenectomy. *Arch. Surg. 104*:587-93, 1972.